Efficiency Bounds for Distribution-Free Estimation
from Endogenously Stratified Samples

Summary of Results

Stephen R. Cosslett
Center for Econometrics and Decision Sciences
University of Florida

August 1985
Abstract

This paper considers estimation of the linear regression model and of the binary choice model from endogenously stratified samples (that is, the strata are defined in terms of the dependent variable), in the case where neither the independent variables nor the error term have a distribution function of known parametric form. This includes, as a special case, the truncated regression model with unknown error distribution. For these models, we derive lower bounds on the asymptotic variances for regular distribution-free (or semi-parametric) estimators, both with and without prior knowledge of the population shares of the strata. For comparison, we also give the corresponding results for the case of a known error distribution.
1. Introduction

A distribution-free or semi-parametric estimator is one that estimates not only a vector of unknown parameters but also one or more unknown distribution functions. Typically the unknown distribution is that of the error term in a regression equation, but in the case of endogenously stratified samples the unknown distribution of the exogenous variables also enters the likelihood in a non-trivial way. For the conventional regression model, least squares methods are of course available. But other cases, such as models with discrete and limited dependent variables, require more complicated methods. Examples include the estimators of Manski (1975) and Cosslett (1983) for the discrete choice model, Buckley and James (1979) for the censored regression model, and Powell (1982) for the censored and truncated regression models.

Although much work remains to be done concerning the asymptotic distribution of semi-parametric estimators, it is useful to have a criterion for asymptotic efficiency—a lower bound on the asymptotic variances analogous to the asymptotic Cramér-Rao bound in the fully parametric case. Using theorems due to Begun, Hall, Huang and Wellner (1983), a recent paper (Cosslett, 1984) has obtained such bounds for the binary choice model and the censored regression model. Similar bounds for the regression model with sample selectivity, and also for the binary choice model, have been obtained by Chamberlain (1984) using a slightly different approach. In the present paper, we apply the method used by Cosslett (1984) to estimation from endogenously stratified samples (with two strata, in the present case).

The models considered here are: (a) the linear regression model, with the strata determined according to whether the dependent variable \( y \) is above or below a given
threshold value (which can be taken as zero without loss); and (b) the binary choice
model, with the strata determined by the choice. Endogenous stratification with discrete
choice models is often called choice-based sampling.

The following special cases are also investigated:

(1) Known population shares. The population share, \( Q_i \), of stratum \( i \) is the
probability that a randomly drawn member of the population falls in that stratum. The
shares \( Q \) may or may not be known a priori, and bounds are derived for both cases. At
present, estimators that attain these bounds are not known.

(2) Truncated regression model. If the sampling probability for one of the strata
is zero, we have the important special case of the truncated regression model with
unknown error distribution. Our results here are not directly comparable with the
asymptotic variance of Powell’s estimator, however, because we do not assume that the
error term has a symmetric distribution.

(3) Known parametric distribution of the error term. This corresponds to the
classical parametric estimation problem, where the error term is generally assumed to be
normally distributed, except that the unknown distribution of the exogenous variables
enters the likelihood in a non-trivial way. (The case of a known parametric distribution
of the exogenous variables does not seem to be realistic, and will not be considered here.)
The non-classical maximum likelihood estimator for this case is known: see Cosslett
(1981a, 1981b) and Manski and McFadden (1981) for the discrete choice case, and
McFadden (1979) for a generalization that includes the endogenously stratified linear
regression model. An earlier paper (Cosslett, 1981a) showed that the choice-based
sampling maximum likelihood estimator is asymptotically efficient in the sense that the
asymptotic variance is equal to the asymptotic limit of the Cramér-Rao bound. Comparison of asymptotic variances with the lower bounds derived in the present paper shows that these estimators are, in fact, asymptotically efficient in the sense of Le Cam.

As before, we should caution that it is not known whether lower bounds of this type are sharp: if an estimator fails to attain the lower bound, this suggests but does not necessarily imply that the estimator can be improved.

2. Notation and basic conditions

The notation and general approach are based on Begun et al. (1983), with some modifications described in Cosslett (1984), and will be briefly restated here in the context of endogenous stratification.

The vector of explanatory variables \( z \in Z \subset \mathbb{R}^r \) has the density function \( h \in H \), and the unobserved variable (error term) \( u \in \mathbb{R}^1 \) has the density function \( g \in G \), where \( G \) and \( H \) are sets of density functions satisfying certain regularity conditions and a priori restrictions. The dependent variable \( y \in Y \) is then determined according to a given model in terms of \( z, u \), and an unknown parameter vector \( \theta \in \Theta \subset \mathbb{R}^k \).

Stratum \( s \) \((s = 1, \ldots, S)\) is then the subpopulation defined by \( y \in J(s) \), where \( J(s) \) is some given subset of the range of \( y \). For example, if \( y \) is a continuous random variable we may have \( J(1) = \{ y \geq 0 \} \) and \( J(2) = \{ y < 0 \} \); if \( y \) is a binary discrete variable we may have \( J(1) = \{ y = 1 \} \) and \( J(2) = \{ y = 0 \} \). (The strata are usually mutually exclusive, and are often exhaustive as well, but they do not have to be.) A random
sample of observations \( x = (y, z) \in X_s \) of size \( N_s \) is then drawn from each stratum \( s \).

The subsample sizes \( N_s \) are assumed to be fixed a priori. Let

\[
H_s = \frac{N_s}{N}
\]

where \( N = \sum_s N_s \) is the total sample size. The asymptotic limit is \( N \to \infty \) with the ratios \( H_s \) held fixed. The estimation problem is to estimate \( \theta \) from the sample \( \{x_{s,i}\}, i = 1, \ldots, N_s, s = 1, \ldots, S \), without knowledge of \( g \) and \( h \). The case of unknown \( h \) but a known parametric form for \( g \) will also be considered.

Examples of endogenous stratification (but with a parametric form for \( g \)) are given by Hausman and Wise (1982) for the linear regression model, and by Manski and Lerman (1977) for the discrete choice model. A special case is the truncated regression model \( (S = 2, H_1 = 1 \text{ and } H_2 = 0) \), and an example of this (again with a parametric form for \( g \)) is given by Hausman and Wise (1977).

The likelihood function for an endogenously stratified sample is

\[
L(x; \theta, g, h) = \prod_{s=1}^{S} \prod_{i=1}^{N_s} f_s(x_{s,i}; \theta, g, h)
\]

with

\[
f_s(x; \theta, g, h) = \frac{p(y; z, \theta, g) h(z)}{Q_s(\theta, g, h)}
\]

where \( p(y; z, \theta, g) \) is the density function for \( y \) conditional on \( z \), as specified by the parametric model. The population share \( Q_s \) is given by

\[
Q_s = \int dz \, h(z) P_s(z, \theta, g)
\]

where
\[ P_s(z, \Theta, g) = \int dy 1(y \in J(s)) p(y; z, \Theta, g) \] (2.4)

is the probability of being in stratum \( s \) conditional on \( z \). Here \( f_s, g \) and \( h \) are densities with respect to sigma-finite measures \( \mu(s), \nu, \) and \( \omega, \) respectively, on some measurable spaces. In the present case \( \nu \) and \( \omega \) will be Lebesgue measure. (If some of the explanatory variables are discrete, then \( \omega \) is the product of Lebesgue measure and counting measure.) Let \( L^2(\mu(s)) = L^2(X_s, \mu(s)), \ L^2(\nu) = L^2(R, \nu) \) and \( L^2(\omega) = L^2(Z, \omega) \) be the corresponding \( L^2 \)-spaces of square-integrable functions, so that \( f_s^{1/2} \in L^2(\mu(s)), \ g^{1/2} \in L^2(\nu) \) and \( h^{1/2} \in L^2(\omega) \). Inner products and norms are denoted by \( \langle \cdot, \cdot \rangle \) and \(|\cdot|\), with subscripts \( \mu(s), \nu \) and \( \omega \) to distinguish the different \( L^2 \)-spaces. We also define

\[ |f^{1/2}|_{\mu}^2 = \sum_{s=1}^{S} H_s |f^{1/2}|_{\mu(s)}^2 \]

and similarly for the scalar product \( \langle \cdot, \cdot \rangle_{\mu} \), where \( f = (f_1, \ldots, f_S) \) and \( f^{1/2} = (f_1^{1/2}, \ldots, f_S^{1/2}) \). Convergence of the sequences \( \{f_n\}, \{g_n\} \) and \( \{h_n\} \) is understood to mean convergence of their square roots in the corresponding \( L^2 \)-spaces. Convergence of sequences \( \{\Theta_n\} \in \Theta \) is with respect to the Euclidean norm \(|\cdot|\) for \( R^k \).

A sequence \( \{ (\Theta_n, g_n, h_n) \} \) with \( \Theta_n \in \Theta, \ g_n \in G \) and \( h_n \in H \) for all \( n \) \((n = 1, 2, \ldots)\) that converges to \( (\Theta, g, h) \), with \( \Theta \in \Theta, \ g \in G \) and \( h \in H \), is said to have “direction of approach” \( (\zeta, \beta, \gamma) \) if
\[ |n^{\frac{1}{2}}(\theta_n - \theta) - \zeta| \to 0 \]
\[ |n^{\frac{1}{2}}(g_n^{\frac{1}{2}} - g^{\frac{1}{2}}) - \beta|_\nu \to 0 \]  \hspace{1cm} (2.5)
\[ |n^{\frac{1}{2}}(h_n^{\frac{1}{2}} - h^{\frac{1}{2}}) - \gamma|_\omega \to 0 \]

as \( n \to \infty \) for some \( \zeta \in \mathbb{R}^k, \beta \in L^2(\nu) \) and \( \gamma \in L^2(\omega) \). Let \( B_g \subset L^2(\nu) \) be the set of all such limit functions \( \beta \) for a given \( g \). It is closed and orthogonal to \( g^{\frac{1}{2}} \). The set \( B_h \subset L^2(\omega) \) is defined similarly. Let \( C(\theta, g, h) \) be the set of all such sequences \( \{((\theta_n, g_n, h_n))\} \).

A regular estimator of \( \theta \) at \((\theta, g, h)\) may be defined as follows. Let \( f_n = f(\cdot; \theta_n, g_n, h_n) \) where \( \{((\theta_n, g_n, h_n))\} \in C(\theta, g, h) \). Then the estimator \( \hat{\theta}_n \) is regular if the distribution function of \( n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n) \) converges to a limiting distribution function which depends on \((\theta, g, h)\) but not on the particular sequence \( \{((\theta_n, g_n, h_n))\} \).

The method of Begun et al. requires the following two conditions:

**Condition 1.** \( B_g \times B_h \) is a subspace of \( L^2(\nu) \times L^2(\omega) \).

**Condition 2.** (Hellinger differentiability). For given \((\theta, g, h)\), there exist a (vector) function \( \rho_\theta \in L^2(\mu) \) and bounded linear operators \( A_g : B_g \to L^2(\mu) \) and \( A_h : B_h \to L^2(\mu) \) such that
\[ |n^{\frac{1}{2}}(f_n^{\frac{1}{2}} - f^{\frac{1}{2}}) - \zeta^T \rho_\theta - A_g^T \beta - A_h^T \gamma|_\mu \to 0 \]
as \( n \to \infty \) for any sequence \( \{((\theta_n, g_n, h_n))\} \in C(\theta, g, h) \).
3. The lower bound

Let \( \alpha_v^* \in L^2(\mu) \) be a function such that

\[
|\rho_{\theta,v} - \alpha_v^*|_\mu = \inf_{\beta,\gamma} |\rho_{\theta,v} - A_g \beta - A_h \gamma|_\mu
\]

(\( v = 1, \ldots, r \)). Then the lower bound on the asymptotic variance matrix of a regular consistent estimator \( \hat{\theta} \) is the inverse of the matrix

\[
I_* = 4 <(\rho_0 - \alpha^*), (\rho_0 - \alpha^*)^T>_\mu
\]

provided this inverse exists. A few obvious changes in the proof given by Begun et al. (1983) allow it to be extended to the present case of \( S \) independent samples, each with its own likelihood function as expressed by eq. (2.1).

Let \( A = \{A_g \beta + A_h \gamma | \beta \in B_g, \gamma \in B_h \} \), and let \( \bar{A} \) be its closure. By the classical projection theorem for Hilbert space, \( \alpha_v^* \) is uniquely characterized by the conditions

\[
\alpha_v^* \in \bar{A}
\]

and

\[
<(\rho_0 - \alpha^*), (A_g \beta + A_h \gamma)>_\mu = 0
\]

for all \( \beta \in B_g \) and \( \gamma \in B_h \). One method of finding \( \alpha^* \) is therefore as follows: first “solve” the minimization problem (3.1) by the variational method, yielding formal solutions \( \beta_v^* \) and \( \gamma_v^* \) (not necessarily in \( B_g \) and \( B_h \)); then verify that \( \alpha_v^* = A_g \beta_v^* + A_h \gamma_v^* \) satisfies conditions (3.3) and (3.4). (Of course, if the conditions did not hold then this method would not work.)
In the cases considered here, \( \alpha^*_v \) is not in fact an element of \( A \), so the main problem is verification of (3.3). One has to construct a sequence \( \{ (\beta_{v,m}, \gamma_{v,m}) \} \) with \( \beta_{v,m} \in B_g \) and \( \gamma_{v,m} \in B_h \) such that \( |\alpha^*_v - A_g \beta_{v,m} - A_h \gamma_{v,m}|_\mu \to 0 \) as \( m \to \infty \). A method of constructing such a sequence is given in Cosslett (1984) in the case of the tobit model, and can be applied here also, but we shall not present the details in the present paper.

When the population shares \( Q_x \) are known a priori, the directions of approach \( \beta \) and \( \gamma \) have to satisfy an inhomogeneous constraint of the form

\[
C_g \beta + C_h \gamma = c
\]  

(3.5)

for some \( C_g : L^2(\nu) \to \mathbb{R} \) and \( C_h : L^2(\omega) \to \mathbb{R} \). At first sight this presents a problem, because Condition 1 no longer holds, so the projection theorem leading to eq. (3.4) cannot be used. Let \( B \) be the set of \( (\beta, \gamma) \) satisfying all conditions including eq. (3.5), and let \( B^0 \) be the same except that eq. (3.5) is replaced by the corresponding homogenous constraint

\[
C_g \beta + C_h \gamma = 0.
\]

Then all we have to do is rewrite the right-hand side of eq. (3.1) as

\[
\inf |(\rho_{0,v} - A_g \beta^c - A_h \gamma^c) - A_g \beta - A_h \gamma|_\mu
\]

where the infimum is taken over \( (\beta, \gamma) \in B^0 \), which is a subspace of \( L^2(\nu) \times L^2(\omega) \), and where \( (\beta^c, \gamma^c) \) is any element of \( B \). The only resulting change is that the orthogonality condition (3.4) is to hold for all \( (\beta, \gamma) \in B^0 \), rather than \( B \).
4. Results for the endogenously stratified regression model

The underlying regression model is

\[ y = z \cdot \theta + u \]

where the error terms \( u \) are i.i.d with density function \( g \). The strata are defined with respect to some (known) threshold value of \( y \), which we may take to be zero without loss, so that \( J(1) = \{ y \geq 0 \} \) and \( J(2) = \{ y < 0 \} \). The components of \( f \) are then

\[
f_s(x; \theta, g, h) = \frac{g(y - z \cdot \theta) h(z)}{Q_s(\theta, g, h)} \tag{4.1}\]

with

\[
Q_s(\theta, g, h) = \int dz \, h(z) [1 - G(-z \cdot \theta)] \tag{4.2}\]

where \( G \) is the distribution function corresponding to \( g \), and \( Q_2 = 1 - Q_1 \). The norm \( |\cdot|_\mu \)
for a function \( \phi(y, z) = (\phi_1(y, z), \phi_2(y, z)) \) is then

\[
|\phi|_\mu^2 = \int dz \left\{ H_1 \int_0^\infty dy [\phi_1(y, z)]^2 + H_2 \int_{-\infty}^0 dy [\phi_2(y, z)]^2 \right\} \tag{4.3}\]

Expressions for the derivatives \( \rho_\theta, A_g \beta \) and \( A_h \gamma \), and their norms and scalar products, are given in an appendix which is available from the author on request. The variational equations for \( \beta \) and \( \gamma \), and their solutions, are also given in the appendix.

Let \( H_s \) be the marginal distribution function of \( -z \cdot \theta \) (where \( \theta \) is the “true” parameter vector), and let \( h_s \) be the corresponding density function. We define

\[
\overline{G}(u) = \frac{H_1}{Q_1} [1 - G(u)] + \frac{H_2}{Q_2} G(u) \tag{4.4}\]

\[
\overline{H}_s(u) = \frac{H_1}{Q_1} H_s(u) + \frac{H_2}{Q_2} [1 - H_s(u)] \tag{4.5}\]

and the weighted-average expectation operator
\[ E(\cdot; u) = [\overline{H}_*(u)]^{-1} \left\{ \frac{H_1}{Q_1} H_*(u) E(\cdot | -z \cdot \theta < u) + \frac{H_2}{Q_2} [1 - H_*(u)] E(\cdot | -z \cdot \theta > u) \right\} \] (4.6)

Then the asymptotic lower bounds are given by the inverses of the following matrices:

(a) \(g\) unknown, \(Q\) unknown

\[ I_* = \int_{u} \left[ \frac{G(u)}{g(u)} \right]^2 \left[ \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) \right]^2 \overline{H}_*(u) \cdot [E(z z^T; u) - E(z; u) E(z^T; u)] \]

\[ - \int_{u} \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) E(z; u) \cdot \int_{u} \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) E(z^T; u) \]

\[ \cdot \left\{ \int_{u} \frac{g(u)}{[G(u)]^2} \left[ 1 - \frac{1}{\overline{H}_*(u)} \right] \right\}^{-1} \] (4.7)

(b) \(g\) unknown, \(Q\) known

\[ I_* = \int_{u} \left[ \frac{G(u)}{g(u)} \right]^2 \left[ \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) \right]^2 \overline{H}_*(u) \cdot [E(z z^T; u) - E(z; u) E(z^T; u)] \]

\[ + \int_{u} \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) E(z; u) \cdot \int_{u} \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) E(z^T; u) \]

\[ \cdot \left\{ \int_{u} \frac{g(u)}{[G(u)]^2} \left[ 1 - \frac{1}{\overline{H}_*(u)} - \left( \frac{H_1}{Q_1} + \frac{H_2}{Q_2} - \frac{H_1 H_2}{Q_1 Q_2} \right)^{-1} \right] \right\}^{-1} \] (4.8)

This differs from (4.7) only in the denominator (and sign) of the last term.

(c) \(g\) known, \(Q\) unknown

\[ I_* = \int_{u} \left[ \frac{G(u)}{g(u)} \right]^2 \left[ \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) \right]^2 \overline{H}_*(u) E(z z^T; u) \]
\[-\left(\frac{H_1}{Q_1} - \frac{H_2}{Q_2}\right)^2 \int du \, h_*(u) \frac{g(u)}{G(u)} \, E(z \mid -z \cdot \theta = u) \cdot \left[ \int du \, h_*(u) \frac{g(u)}{G(u)} \, E(z^T \mid -z \cdot \theta = u) \right]^{-1}\]

\[= \left\{ \frac{Q_1^2}{H_1} + \frac{Q_2^2}{H_2} - \left[ \int du \, h_*(u) \frac{g(u)}{G(u)} \right]^{-1} \right\}^{-1} \]

(4.9)

(d) \(g\) known, \(Q\) known

\[I_* = \int du \frac{[\tilde{G}(u)]^2}{g(u)} \left[ \frac{d}{du} \left( \frac{g(u)}{G(u)} \right) \right]^2 \tilde{H}_*(u) \tilde{E}(z z^T; u)\]

\[+ \left(\frac{H_1}{Q_1} - \frac{H_2}{Q_2}\right)^2 \int du \, h_*(u) \frac{g(u)}{G(u)} \, E(z \mid -z \cdot \theta = u) \cdot \left[ \int du \, h_*(u) \frac{g(u)}{G(u)} \right]^{-1}\]

Again, this differs from the corresponding expression for unknown \(Q\), (4.9), only in the denominator (and sign) of the last term.

In the special case \(H_1 = Q_1, \, H_2 = Q_2\) (i.e., the sampling proportions are the same as the population proportions), both the numerator and the denominator of the second term vanish in each of eqs. (4.7)-(4.10), but the correct expression can be obtained by taking the limit \((H_1 - Q_1) \to 0\) in the usual way.

First, we note that, in general, (4.7) and (4.8) are different from (4.9) and (4.10). Thus what Begun et al. refer to as “adaptive estimation” appears not to be possible in general – that is, unlike the case of estimation from a random sample, one cannot attain the same efficiency when \(g\) is unknown as one can when \(g\) is known. [It might be better to use the term “adaptive estimation” to denote attainment of the actual lower bound (4.7)]
or (4.8), when $g$ is unknown.] Numerical computations indicate that the loss of efficiency when $g$ is unknown can be substantial.

Secondly, we note that the lower bounds given by (4.9) and (4.10) are equal to the asymptotic variance of the estimator proposed by McFadden (1979) for known $g$, when specialized to the present cases, and therefore that estimator is asymptotically efficient in these cases. The conditional maximum likelihood estimator used by Hausman and Wise (1982) does not attain the lower bound; heuristically, the reason is that their estimator does not use the information that the density of $z$ in each stratum is generated by the same population density $h(z)$.

5. Results for the truncated regression model

The truncated regression model can be considered as a special case of endogenous stratification in which $H_2 = 0$, i.e., subjects with $y < 0$ are dropped from the sample. Since this is an important case, the results are presented separately.

(a) $g$ unknown, $Q$ unknown

$$I_* = \frac{1}{Q_1} \int du \frac{[1 - G(u)]^2}{g(u)} \left[ \frac{d}{du} \left( \frac{g(u)}{1 - G(u)} \right) \right]^2 H*(u) \text{var}[z \mid z \cdot \theta < u]$$ (5.1)

It may perhaps be surprising that there is any finite bound at all in this case, since so little information is available.

(b) $g$ unknown, $Q$ known

This differs from (5.1) by the additional term

$$\frac{1}{Q_1} \int du \frac{d}{du} \left( \frac{g(u)}{1 - G(u)} \right) E[z \mid z \cdot \theta < u] \cdot \int du \frac{d}{du} \left( \frac{g(u)}{1 - G(u)} \right) E[z^T \mid z \cdot \theta < u]$$
\[
\left\{ \int du \frac{g(u)}{[1-G(u)]^2} \left( \frac{1}{H_s(u)} - 1 \right) \right\}^{-1}
\]  

(5.2)

If the integral in the denominator diverges, then knowledge of \( Q \) does not improve the bound.

The density \( f_1(x; \theta, g, h) \) can be rewritten as

\[
f_1 = \frac{g(y - z \cdot \theta)}{1 - G(-z \cdot \theta)} h^*(z)
\]

(5.3)

where \( h^* \) is the density function of \( z \) conditional on the subject being in the observed stratum. Thus when \( g \) is known (and \( Q \) unknown), one has a classical parametric likelihood function \( g(y - z \cdot \theta) / [1 - G(-z \cdot \theta)] \) for \( \theta \), and the classical maximum likelihood estimator is asymptotically efficient.

6. Results for the endogenously stratified binary choice model

In this model, the choices \( y = 0 \) or \( 1 \) are generated by the sign of a latent (unobserved) variable \( y^* \):

\[
y^* = z \cdot \theta + u
\]

\[
y = \begin{cases} 
0 & \text{if } y^* < 0 \\
1 & \text{if } y^* \geq 0
\end{cases}
\]

where, as before, the error terms are i.i.d. with density function \( g \). The strata just correspond to the values of \( y \), i.e., \( J(1) = \{ y = 1 \} \) and \( J(2) = \{ y = 0 \} \). The components of \( f \) are then
with $Q$ defined as before (eq. 4.2). The norm $|\cdot|$ for a function $\phi(z) = (\phi_1(z), \phi_2(z))$ is then

$$|\phi|_\mu^2 = \int dz \{H_1[\phi_1(z)]^2 + H_2[\phi_2(z)]^2\}$$

Using the notation of eq. (4.4), the asymptotic lower bounds are given by the inverses of the following matrices:

(a) $g$ unknown, $Q$ unknown

$$I_* = \frac{H_1H_2}{Q_1Q_2} \int du h_*(u) \frac{[g(u)]^2}{G(u)G(u)[1 - G(u)]} \text{var}[z \mid -z \cdot \theta = u]$$

(b) $g$ unknown, $Q$ known

This leads to the same bound, (6.3), so knowledge of $Q$ is not useful here.

(c) $g$ known, $Q$ unknown

$$I_* = \frac{H_1H_2}{Q_1Q_2} \int du h_*(u) \frac{[g(u)]^2}{G(u)G(u)[1 - G(u)]} E[z z^T \mid -z \cdot \theta = u]$$

$$- \left( \frac{H_1}{Q_1} - \frac{H_2}{Q_2} \right)^2 \int du h_*(u) \frac{g(u)}{G(u)} E[z \mid -z \cdot \theta = u].$$

$$\cdot \int du h_*(u) \frac{g(u)}{G(u)} E[z z^T \mid -z \cdot \theta = u] \cdot \left\{ \frac{Q_1^2}{H_1} + \frac{Q_2^2}{H_2} - \int du \frac{h_*(u)}{G(u)} \right\}^{-1}$$

(d) $g$ known, $Q$ known

$$I_* = \frac{H_1H_2}{Q_1Q_2} \int du h_*(u) \frac{[g(u)]^2}{G(u)G(u)[1 - G(u)]} E[z z^T \mid -z \cdot \theta = u]$$
\[
+ \left( \frac{H_1}{Q_1} - \frac{H_2}{Q_2} \right)^2 \int du \ h_s(u) \ \frac{g(u)}{G(u)} \ E[z \mid -z \cdot \theta = u].
\]

\[
\cdot \int du \ h_s(u) \ \frac{g(u)}{G(u)} \ E[z^T \mid -z \cdot \theta = u] \cdot \left\{ \int du \ \frac{h_s(u)}{G(u)} - 1 \right\}^{-1}
\]

(6.5)

In general, knowledge of the parametric form of \(g\) leads to an improvement (assuming the bounds can be attained). However, if the distribution of \(z\) is multivariate normal then the bounds are the same (since \(\theta\) can only be estimated up to overall scale and location parameters). This was previously found in the case of random sampling of the binary choice model (Cosslett, 1984).

The bounds given by (6.4) and (6.5) are equal to the asymptotic variances of the corresponding nonclassical maximum likelihood estimators (see equations 2.45 and 2.78 in Cosslett, 1981b), which are therefore asymptotically efficient.
References


